CRC 336

AD A046773

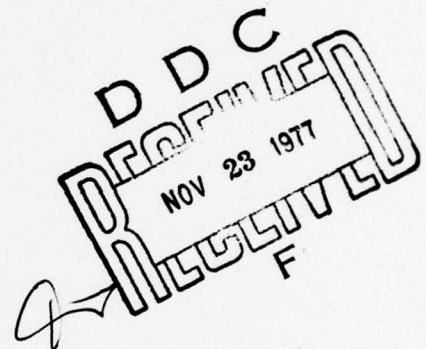# A METHOD TO CORRECT CORRELATION COEFFICIENTS FOR THE EFFECTS OF MULTIPLE CURTAILMENT

CENTER FOR NAVAL ANALYSES

1401 Wilson Boulevard
Arlington, Virginia 22209

Marine Corps Operations Analysis Group

By: Thomas L. Mifflin, Stephen M. Verna

D D C
RECEIVED
NOV 23 1977
F

August 1977

AD No.
DDC FILE COPY

02 033600.00

REPORT DOCUMENTATION PAGE

**READ INSTRUCTIONS
BEFORE COMPLETING FORM**

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
CRC-336		

4. TITLE (and Subtitle)

A Method to Correct Correlation Coefficients for the Effects of Multiple Curtailment

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

Thomas L. Mifflin and Stephen M. Verna

8. CONTRACT OR GRANT NUMBER(s)

N00014-76-C-0746

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Center for Naval Analyses
1401 Wilson Boulevard
Arlington, Virginia 22209

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Office of Naval Research
Department of the Navy
Arlington, Virginia 22217

12. REPORT DATE

August 1977

13. NUMBER OF PAGES

32

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

Deputy Chief of Staff(RD&S)
Headquarters, Marine Corps
Washington, D.C. 20380

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

D D C
RECEIVED
NOV 23 1977
F

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This Research Contribution does not necessarily represent the opinion of the Department of the Navy.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

coefficients, computer programs, correlation techniques, corrections, mathematical logic, validations

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report examines a general method to estimate the correlation of variables in a large, diverse population from a smaller, selected population in which the ranges of the variables have been restricted. All correlation coefficients are corrected simultaneously, so that an entire correlation matrix can be corrected. Alternate methods are compared. The effectiveness of the method is tested, and an example is given.

180

DD FORM 1473 1 JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

HO7 327.

4B
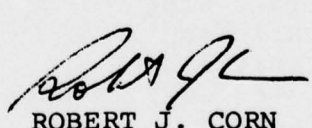
19 October 1977

MEMORANDUM FOR DISTRIBUTION LIST

Subj: Center for Naval Analyses Research Contribution 336

Encl: (1)   CRC 336, "A Method to Correct Correlation Coefficients
              for the Effects of Multiple Curtailment," by Thomas L.
              Mifflin and Stephen M. Verna, August 1977

1. Enclosure (1) is forwarded as a matter of possible interest.

2. This Research Contribution examines a problem frequently
encountered in correlation analyses -- that of range restriction,
or curtailment. The results may be useful in recognizing and
correcting for such restrictions, as well as for interpreting and
understanding correlation analyses in general.

3. Research Contributions are distributed for their potential value
in other studies and analyses. They do not necessarily represent the
opinion of the U.S. Marine Corps or the Department of the Navy.

4. The enclosure is approved for public release.

ROBERT J. CORN
Director
Marine Corps Operations
    Analysis Group

Subj: Center for Naval Analyses Research Contribution 336

DISTRIBUTION:

HQMC
   DC/S for Research, Development and Studies (2)
   DC/S for Manpower (10)
   DC/S for Operations and Training (10)
   Director of Marine Corps Reserves (2)
OPNAV
   OP-01, DCNO, Manpower (2)
   OP-59, Aviation Manpower and Training (2)
   OP-099, Naval Education and Training (2)
   OP-09BH, Naval History
Fleet Training Command, Atlantic
Fleet Training Command, Pacific
Fleet Marine Force Command, Atlantic
Fleet Marine Force Command, Pacific
Landing Force Training Command, Atlantic
Landing Force Training Command, Pacific
Fleet Aviation Specialized Operational Training Group, Atlantic
Fleet Aviation Specialized Operational Training Group, Pacific
First Marine Division
Second Marine Division
Third Marine Division
Fourth Marine Division
Force Troops, Atlantic/Second Force Service Support Group
Force Troops, Pacific
First Marine Aircraft Wing
Second Marine Aircraft Wing
Third Marine Aircraft Wing
Fourth Marine Aircraft Wing
Ass't Secretary of the Navy, Manpower & Reserve Affairs
Office of Naval Research
Chief of Naval Personnel
National Defense University
Armed Forces Staff College
Human Resource Management Center, Atlantic
Human Resource Management Center, Pacific
Human Resource Management Center, London
Manpower and Material Analysis Center, Atlantic
Manpower and Material Analysis Center, Pacific
U.S. Naval Academy, Nimitz Library
Personnel Program Support Activity
Recruiting Command
Personnel Research and Development Center
Chief of Naval Education and Training

DISTRIBUTION (Continued)

Chief of Naval Air Training
Chief of Naval Technical Training
Naval Aviation Schools Command
Naval Air Technical Training Center, Lakehurst
Naval Air Technical Training Center, Millington
Fleet Combat Training Center, Atlantic
Fleet Combat Training Center, Pacific
Naval Postgraduate School (2)
Naval War College
Human Resource Management School
Marine Corps Recruit Depot, Paris Island
Marine Corps Recruit Depot, San Diego
Marine Corps Development and Education Command
Marine Corps Base, Camp Lejeune
Marine Corps Base, Camp Pendleton
Marine Corps Base, Camp Smith
Marine Corps Base, Camp Butler
Marine Corps Base, Twentynine Palms
Director, Defense Research and Engineering
Asst. SecDef, Planning Analysis and Evaluation (2)
Defense Documentation Center (12)
Department of the Army (Adj Gen'l) (6)
Department of the Air Force (AFXODD)
Human Resource Research Organization

TABLE OF CONTENTS

SUMMARY

This report examines a method to correct for the effect of range restriction on correlation coefficients. Often, it is necessary to estimate the correlation of two variables in a large, diverse population from a smaller, selected population in which the ranges of the variables have been restricted. Range restriction (sometimes called curtailment) occurs whenever there is a real change of variance in a particular variable in the selected population. A direct calculation of the correlation coefficient for the larger group from the data sample of the smaller group is misleading. To more accurately estimate the true correlation, methods of correcting for range restriction have been developed.

The method described in this report is one of the most general. It handles the case in which the sample population has been directly restricted in several variables. All coefficients are corrected simultaneously, so that an entire correlation matrix can be corrected. This report gives the correction equations, and the assumptions needed to derive those equations. A FORTRAN program, used to compute the corrected correlation coefficients, is given in appendix A; an APL program in appendix B.

The correction method is illustrated using the aptitude test scores of Marines selected for formal training and similar test data for 60,000 FY-1975 Marine recruits. An empirical test for the effectiveness of the equations is given. In addition, another technique for correcting for range restriction is discussed and compared. In general, the method documented in this report appeared superior to the other examined.

INTRODUCTION

Range restriction is a problem that is often encountered in correlation analysis. Recall that the correlation coefficient r quantifies the extent that two variables covary in a particular population. It is misleading to speak of the correlation between two variables without specifying the sample population, since, generally, the size of r is related to the ranges of the correlated variables in the measured population. Usually, the correlation coefficient computed from a population in which the ranges of the variables have been restricted will be smaller than the r computed from a broader, unrestricted population. Since it is often desirable to estimate the correlation of two variables in a large population from data obtained from a more restricted population, it is necessary to correct the correlation coefficients computed in the smaller population for the effects of range restriction.

Suppose, for example, that a group of people are given intelligence test A, and that only those who score above 90 are administered test B. Thus, the results of test A are used to restrict the group who take test B. It is often of interest to find how well test A predicts performance on test B. That is, what is the correlation between tests A and B for the total group? Because the group that took both tests A and B was restricted on the basis of performance on test A, that question cannot be answered by direct calculation of the correlation coefficient. The problem would be even more complex if several tests were the basis of restriction. This report examines a method of correcting for range restriction that can handle the problem of multiple curtailment. Another, simpler method will also be discussed and compared.

CORRECTION FOR MULTIPLE CURTAILMENT

The problem of range restriction often occurs when the assignment of personnel to jobs or schools is based on test performance. The validity of a test is measured by how accurately it predicts the later performance of a member of the general, or unrestricted, population. A direct measurement of a test's validity is impossible if that same test is used for personnel selection, since performance measures exist only for the subset of the general population selected for a job. In order to estimate accurately the validity of a test in the general population, the correlation coefficients calculated from the selected population must be corrected for the effects of range restriction.

A recent study of Marine Corps school performance (reference 1) provides an example of the problem of range restriction, and will be used throughout this report to illustrate how the problem may be solved. At the time of the study, Army Classification Battery (ACB-61) of 11 subtests was administered to all Marine recruits at the recruit depot. Although the range of scores varied slightly among the different tests, the approximate range was from 50 to 160, and all the test scores were approximately normally distributed. A number of composite scores were computed from linear combinations of the subtests. Table 1 lists the subtests and composites.

After recruit training, recruits were selected for assignment to jobs and formal training schools based on their ACB-61 scores. For example, to be admitted into the Sea Duty Indoctrination School, a Marine had to score 90 or above on his General Technical (GT) test. By thus restricting the range of GT scores, the variance of the GT scores in the selected school population was reduced. At the end of school training, a recruit was assigned a final course grade (FCG) ranging from 0 to 100.

In order to evaluate how well GT predicts a recruit's performance in any course, the correlation coefficient between GT and FCG may be computed. Figure 1 illustrates the problem of computing that correlation when the ranges of the students' GT scores have been restricted. In a typical school, if there were no entrance requirements, a scatterplot of the students' FCG versus GT scores might look like figure 1a. In contrast, figure 1b shows what the same scatterplot would look like if a GT score of 90 or above were required for admission into the school. In general, the correlation coefficient of the restricted school population is less than that of the unrestricted, general population. When several different variables are used as the basis of restriction, the problem is even more complex.

Range restriction can occur from above as well as below. For example, if all the recruits with high GT scores are selected for more demanding schools, few will be available for assignment to "easier" schools, thus the range of GT scores of students in those schools will have been restricted from above. The curtailment of GT scores from above will, in general, reduce the variance of the distribution of scores; the correlation between GT and FCG will also be reduced.

-3-

TABLE 1

ACB-61 SUBTESTS

Subtest	Abbreviation
Verbal	VE
Arithmetic reasoning	AR
Pattern analysis	PA
Classification inventory	CI
Mechanical aptitude	MA
Army clerical speed	ACS
Army radio code	ARC
General information test	GIT
Shop mechanics	SM
Automotive information	AI
Electronics information	ELI

ACB-61 COMPOSITE TESTS

Composite		Abbreviation
Infantry Combat	$\dfrac{AR + 2\,CI}{3}$	IN
Armor, Artillery, Combat Engineers	$\dfrac{GIT + AI}{2}$	AE
Electronics	$\dfrac{MA + 2\,ELI}{3}$	EL
General Maintenance	$\dfrac{PA + 2\,SM}{3}$	GM
Mechanical Maintenance	$\dfrac{SM + AI}{2}$	MM
Clerical	$\dfrac{VE + ACS}{2}$	CL
General Technical	$\dfrac{VE + AR}{2}$	GT
General Classification Test	$\dfrac{VE + AR + PA}{3}$	GCT

-4-
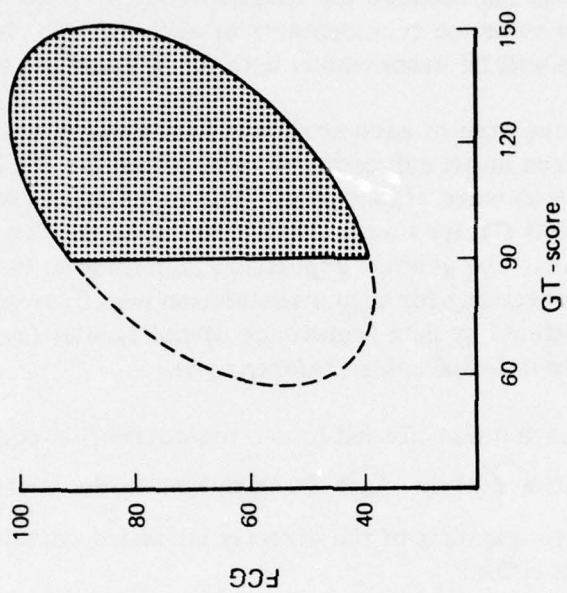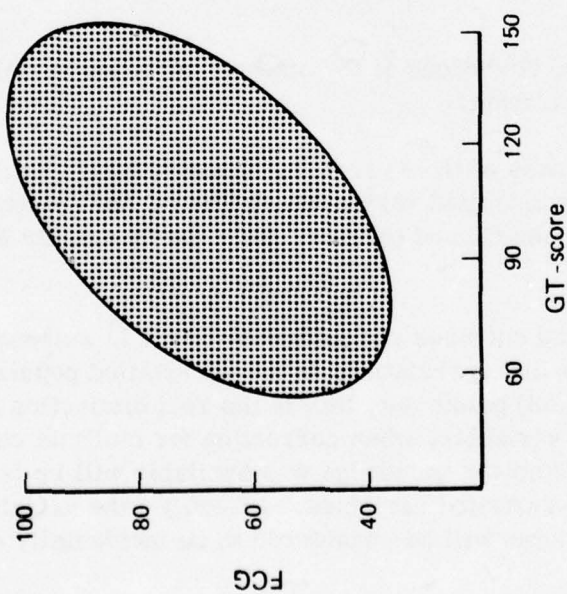
FIG. 1: AN EXAMPLE OF RANGE RESTRICTION

One of the objectives of the school performance study (reference 1) was to determine how well the ACB-61 subtests and composites predicted each recruit's final course grade. To accomplish this, the correlation matrix of all of the tests and the FCG had to be corrected for range restriction because the ranges of one or more of the scores had been directly restricted by the entrance requirements of each school. In the remainder of this report, school population will be synonymous with the restricted, or curtailed, population.

The data available consisted of each student's FCG and his score on each subtest and composite. The scores of all subtests and composites for 60,000 FY-1975 Marine nonprior-service accessions were also known. The correlation coefficients computed from the data on the 60,000 Marines were considered to be the "true" correlations in the general population (i.e., the general population is defined to be all Marine recruits). The technique used for correcting for range restriction was first developed by Pearson (reference 2) and later refined by Burt (reference 3) and Lawley (reference 4). The notation used in this study is Gulliksen's (reference 5).

The following information was needed to use the correction equations:

(1) The correlation matrix of all the variables in the restricted population;

(2) The correlation matrix of the directly curtailed variables in the unrestricted population;

(3) The standard deviations of all the variables in the restricted population; and

(4) The standard deviations of the directly curtailed variables in the unrestricted population.

Of the above, in the case of the 84 training schools, the 11 ACB-61 subtests were designated as the directly curtailed variables. Items (1) and (3) were obtained on recruits in formal schools, and items (2) and (4) came from the data of the 60,000 FY-1975 Marine accessions.

Although no school had entrance requirements on all 11 subtests, complete knowledge of the standard deviations and correlations in the uncurtailed population was available for them. As Burt (reference 3) points out, this is the real distinction between the directly and indirectly restricted variables, when correcting for multiple curtailment. Therefore, the variables for which complete knowledge was available will be considered to be explicitly selected or directly curtailed variables. Likewise, the variables for which only incomplete data was available will be considered to be incidentally selected or indirectly curtailed variables.

Assumptions of the Procedure

All variables are assumed to be normally distributed. The variables subject to incidental selection are regarded as being estimated by linear combination of the explicit selection variables. In the school example, this means that

$$FCG = b_1 (VE) + b_2 (AR) + \ldots + b_{11} (ELI).$$

Furthermore, the gross score weights applied to the explicit selection variables (i.e., b_i) are assumed to be the same for the curtailed and uncurtailed population. This assumption in the univariate case (where FCG is estimated by only one test score) would mean that the regression lines in the unrestricted and restricted population would have equal slopes. Also, it is assumed that the errors of estimate (i.e., the differences between the predicted and actual value of the incidentally selected variables) are the same for both the unrestricted and restricted groups. Finally, after the effects of the explicitly selected variables are partialled out, it is assumed that the correlations among the variables subject to incidental selection in the curtailed population are the same as the analogous partial correlations in the uncurtailed population. In a three-variable case, say variables x, y, and z, it is assumed that for constant z the correlation between x and y is the same in both the unrestricted and restricted populations. This last assumption is examined in greater detail in appendix C.

The Correction Equations

The equations used to correct for the effects of range restriction will now be presented in general form, and will be demonstrated by the example mentioned previously.

Suppose that each member of a population P is administered tests V_1, V_2, ..., V_a, and his score is recorded. Furthermore, suppose that a subpopulation Q of P is obtained by requiring that an individual in Q scores in a particular range on tests V_1, V_2, ..., V_a. In the example, P would be the population of all Marine recruits, V_i the 11 ACB-61 subtests, and Q the subpopulation of all Marines admitted into a particular school. Also, suppose the members of Q are administered tests V_{a+1}, V_{a+2}, ..., V_{a+t}, and their scores are recorded. (In the example, this would be the final exam in a particular school.) This would make V_{a+1}, V_{a+2}, ..., V_{a+t} the variables subject to incidental selection. If a sample is taken from the restricted population Q the correlation between tests V_i and V_j, for i, j = 1, 2, ..., a+t, can be estimated. Denote, by $\hat{C}$, the matrix of correlation coefficients calculated from a sample of Q. This would make $\hat{C}$ an (a+t) x (a+t) square matrix. In the example, $\hat{C}$ was the 20 x 20 correlation matrix of the 11 ACB-61 subtests, the 8 composites, and the FCG for any one school. Suppose it is necessary to estimate what the correlation matrix of tests V_1, V_2, ..., V_{a+t} would be if everyone in P had

-7-

also taken tests V_{a+1}, V_{a+2}, ..., V_{a+t}. This correlation matrix, not yet calculated, of all the variables in the unrestricted population will be called $\hat{D}$. Part of $\hat{D}$ can be estimated directly from a sample of P, since everyone in P has taken tests V_1, V_2, ..., V_a. Therefore, the a x a submatrix of $\hat{D}$, consisting of the correlations between the first a variables, will be called $\hat{D}_{aa}$. Still unknown is the correlation submatrix between tests V_1, V_2, ..., V_a and tests V_{a+1}, V_{a+2}, ..., V_{a+t}. Denote this a x t submatrix of $\hat{D}$ by $\hat{D}_{at}$, and let $\hat{D}_{ta}$ be its transpose. Also the correlation submatrix between the incidentally selected variables (that is, variables V_{a+1}, V_{a+2}, ..., V_{a+t}) is still unknown. Denote this t x t matrix by $\hat{D}_{tt}$. Partition $\hat{C}$ in a similar manner. That is, let $\hat{C}_{at}$ be the a x t submatrix of $\hat{C}$ consisting of the correlations of the directly restricted variables and the incidentally selected variables, and let $\hat{C}_{ta}$ be its transpose. Likewise, let $\hat{C}_{tt}$ be the correlation submatrix of the last t variables (that is, the correlation matrix between the incidentally selected variables.

In order to use a multiple curtailment method, it is necessary to convert $\hat{C}$ and $\hat{D}_{aa}$ into variance-covariance matrixes C and D_{aa}, using:

$$\text{Cov}(V_i, V_j) = \sigma_i \cdot \sigma_j \cdot r_{ij},$$

where r_{ij} denotes the correlation of V_i with V_j, $\text{Cov}(V_i, V_j)$ stands for the covariance of V_i and V_j, and σ_i is the standard deviation of V_i. For example, if $\hat{C}$ is to be transformed into C, then σ_i estimates the standard deviations of V_i in the restricted population. Likewise, if the D matrix is being calculated, σ_i estimates the standard deviation in the unrestricted population. Then, the corrected variance-covariance submatrixes are calculated as follows:

$$D_{ta} = C_{ta} C_{aa}^{-1} D_{aa} ,$$

and

$$D = C_{tt} + C_{ta} C_{aa}^{-1} (D_{at} - C_{at})$$

(see reference 5).

-8-

The diagonal of the D_{tt} matrix consists of the variance of the directly restricted variables in the unrestricted population. Therefore (after taking the square root of these variances) to estimate the σ_i of the incidentally selected variables in the unrestricted population, D can be converted into the correlation matrix D by:

$$r_{ij} = Cov(V_i, V_j)/\sigma_i \cdot \sigma_j \quad i, j = 1, 2, \ldots, a+t.$$

In the example, only the 11 x 11 matrix of correlations of the ACB-61 subtests were included in $\hat{C}_{aa}$. This is because the composite tests are linear combinations of the ACB-61 subtests; and, since C_{aa}^{-1} must be computed in the correction equation, only linearly independent variables can be included in the set of directly curtailed variables. Figures 2 and 3 show the $\hat{C}$ and $\hat{D}$ matrixes in PIBAD (an administrative school) divided into submatrixes.

$\hat{C}_{aa}$ $\hat{C}_{at}$ $\hat{C}_{ta}$ $\hat{C}_{tt}$

	VE	AR	PA	CI	MA	ACS	ARC	GIT	SM	AI	ELI	IN	AE	EL	GM	MM	CL	GT	GCT	FCG
VE	100	60	42	42	51	38	29	66	53	36	40	54	59	49	56	49	85	90	83	41
AF	60	100	51	35	49	51	39	53	58	33	34	59	49	43	57	46	67	89	85	46
PA	42	51	100	26	47	42	37	39	49	37	45	37	44	52	76	48	51	52	78	41
CI	42	35	26	100	43	39	28	37	33	24	24	96	35	33	35	31	44	43	41	20
MA	51	49	47	43	100	44	33	59	64	54	46	51	65	69	66	65	57	56	60	31
ACS	38	51	42	39	44	100	46	39	43	24	25	49	36	34	49	37	81	50	53	37
APC	29	39	37	24	33	46	100	31	31	13	16	35	25	24	38	24	44	38	42	33
GIT	66	53	39	37	59	39	31	100	61	52	47	48	66	57	61	63	64	67	55	37
SM	51	50	49	33	64	43	31	61	100	64	55	53	72	63	94	91	58	54	62	29
AI	35	33	37	24	54	24	13	52	64	100	55	30	86	62	63	90	36	38	43	18
ELI	43	34	45	24	46	25	16	47	53	55	100	31	50	96	58	60	40	41	48	30
IN	54	57	37	96	51	49	35	48	43	31	31	100	45	41	47	41	62	63	61	30
AE	53	49	44	35	65	36	25	66	72	86	54	45	100	68	71	87	58	61	62	32
EL	49	43	52	33	69	34	24	57	63	62	96	41	68	100	68	69	51	58	50	34
GM	56	57	76	35	66	49	38	61	94	63	54	47	71	68	100	87	64	63	77	38
MM	49	46	48	31	65	37	24	63	91	90	60	41	87	69	87	100	52	53	58	26
CL	85	67	51	44	57	81	44	64	58	36	40	62	58	51	64	52	100	86	83	40
GT	90	89	52	43	56	50	38	67	54	38	41	63	61	58	63	53	86	100	94	48
GCT	83	85	78	41	60	53	42	55	62	43	48	61	62	50	77	58	83	94	100	52
FCG	41	46	41	20	31	37	33	37	29	18	30	30	32	34	38	26	40	48	52	100

FIG. 2: Ĉ MATRIX FOR PIBAD

-10-

	VE	AR	PA	CI	MA	ACS	ARC	GIT	SM	AI	ELI	IN	AE	EL	GM	MM	CL	GT	GCT	FCG
VE	100	66	49	50	57	42	41	69	59	45	50	61	64	58	63	57	85	91	85	47
AE	66	100	57	44	53	52	44	58	53	43	45	66	57	52	62	53	70	91	84	51
PA	49	57	100	35	51	46	39	47	50	40	48	46	51	54	78	51	57	58	81	46
CI	57	44	35	100	46	42	29	51	45	36	33	96	49	41	47	45	55	52	51	26
MA	57	53	51	46	100	45	38	60	66	59	51	54	67	73	69	64	61	60	63	35
ACS	42	52	46	42	45	100	46	43	44	33	29	5C	43	30	51	42	83	51	55	41
APC	41	44	39	29	35	46	100	37	35	25	26	37	35	33	42	33	52	47	49	37
GIT	69	58	47	51	60	43	37	10C	63	59	53	59	89	51	66	67	67	78	69	42
SM	59	53	5C	45	66	44	35	63	100	67	56	53	73	66	94	91	61	62	64	33
AI	45	43	43	36	59	33	25	59	67	100	57	43	89	64	67	92	47	48	52	26
ELI	50	45	48	33	51	29	26	53	56	57	110	41	62	96	60	62	47	52	56	36
IN	41	66	46	96	54	50	37	59	53	43	41	100	57	70	59	53	66	78	60	37
AE	54	57	51	49	67	43	35	89	73	89	62	57	130	71	74	99	64	67	68	38
EL	58	52	54	41	73	38	33	61	66	64	96	50	71	10C	70	71	57	61	65	40
GM	53	62	78	47	69	51	42	66	94	67	63	58	74	70	100	88	66	69	88	42
MM	57	53	51	45	68	42	33	67	91	92	62	53	89	71	98	100	59	60	63	32
CL	85	70	57	55	61	41	52	67	61	47	47	66	64	57	64	59	100	86	44	52
GT	91	91	58	52	60	51	47	76	62	48	52	70	67	51	69	60	86	120	95	54
SCT	95	8R	81	51	63	55	48	69	64	52	56	68	68	65	80	63	84	95	100	57
FCG	47	51	46	26	35	41	37	42	73	76	36	37	38	40	42	32	52	54	57	1CG

D̂aa   D̂at

D̂ta   D̂tt

FIG. 3: D̂ MATRIX FOR PIBAD

-11-

COMPARISON WITH ANOTHER TECHNIQUE

Another method for correcting for range restriction is given by Thorndike (reference 6). This method was originally developed by Karl Pearson, and is actually just a special case of the multiple curtailment model already discussed. Thorndike's method assumes direct curtailment on only one variable. It has the advantage of being easier and requiring less information than the equations used for multiple direct curtailment. However, it is neither as general nor, as will be shown empirically, as accurate as the multiple direct curtailment model.

When only one variable has been directly curtailed, there are two basic formulas for correcting for range restriction. If variable 3 has been restricted, then R_{12}, defined to be the corrected correlation coefficient between variable 1 and variable 2, is given by:

$$R_{12} = \frac{r_{12} + r_{13}\, r_{23} \left(\dfrac{\hat{S}_3^{\,2}}{S_3^{\,2}} - 1 \right)}{\sqrt{\left(1 + r_{13}^{\,2} \left(\dfrac{\hat{S}_3^{\,2}}{S_3^{\,2}} - 1 \right) \right) \left(1 + r_{23}^{\,2} \left(\dfrac{\hat{S}_3^{\,2}}{S_3^{\,2}} - 1 \right) \right)}}$$

where r_{ij} is the uncorrected correlation coefficient between variable i and variable j, $\hat{S}_3$ the standard deviation of variable 3 in the unrestricted population, and S_3 the standard deviation of variable 3 in the restricted population. When variable 3 and variable 1 are the same, the above equation reduces to:

$$R_{12} = \frac{r_{12} \dfrac{\hat{S}_1}{S_1}}{\sqrt{1 - r_{12}^{\,2} + r_{12}^{\,2} \dfrac{\hat{S}_1^{\,2}}{S_1^{\,2}}}} \, .$$

In order to check and compare the two correction methods, a test was conducted. A correlation matrix of the 11 ACB-61 subtests was calculated for each of 26 Marine schools with sample populations of 225 to 2,400 students.

-12-

A correlation matrix of the same 11 variables was calculated from the sample of 60,000 FY-1975 Marines. As before, this matrix was assumed to represent the "true" correlation matrix in the general population. In order to use the multiple curtailment method, the first seven ACB-61 subtests were arbitrarily designated as the directly restricted variables. That is, the 7 x 7 variance-covariance matrix of VE, AR, PA, CI, MA, ACS, and ARC (computed from the data on the 60,000 Marines) made up the D_{aa}

matrix of the model, while the variance-covariance matrix of all 11 subtests computed from each school's data made up the C matrix. The range correction equations gave an estimate of the 4 x 4 correlation matrix of subtests subject to "incidental" curtailment, and 4 x 7 and 7 x 4 submatrixes between indirectly and directly curtailed variables. Comparing these submatrixes with the corresponding submatrixes of the true correlation matrix gave an indication of the accuracy of the range correction equations. Similarly, assuming direct curtailment only on VE, each school's correlation matrix was corrected for range restriction, using Thorndike's equations.

The matrix of correlations found by using each of the two methods was subtracted from the matrix of true correlations, and the entries of the three difference matrixes were squared and summed. To fairly compare the two correction techniques, the entries in the 7 x 7 submatrix of differences of correlation coefficients of the first seven variables (the variables on which direct curtailment was assumed in the multiple curtailment method) were not squared and summed.

Expressing this in matrix notation, let M(i, j) be the correlation matrix corrected by the multiple curtailment method. Similarly, let S(i, j) represent the matrix corrected by assuming direct curtailment on only a single variable, VE. Let T(i, j) denote the "true" correlation matrix. Then, define the multiple variable index of accuracy (MVA) as:

$$MVA = \sum_{i,j} \left(T(i,j) - M(i,j) \right)^2 \quad ,$$

$$i \text{ or } j > 7$$

and define the single variable index of accuracy (SVA) as:

$$SVA = \sum_{i,j} \left(T(i,j) - S(i,j) \right)^2$$

$$i \text{ or } j > 7 \quad .$$

-13-

Table 2 shows the SVAs and MVAs for each of the 26 Marine Corps schools. Table 3 gives the names and abbreviations of the schools used in table 2. Tables 4 through 9 show the relevant matrixes for one school, AGAFAM.

With respect to the test used in this analysis to check and compare the two correction methods for range restriction, the multiple curtailment method is superior to Thorndike's. This conclusion is based only on empirical evidence. However, in many cases there are theoretical reasons for assuming that more than one variable is directly curtailed. In addition, the multiple curtailment method allows the analyst to use all of the true correlation coefficients available. This last reason is particularly important if one wants to use the corrected correlation coefficient matrix in a multiple regression or factor analysis.

TABLE 2

SVA AND MVA FOR 26 MARINE CORPS SCHOOLS

School	SVA	MVA	SVA/MVA
FOODSER	.1950	.0738	2.6
MP	.5329	.2464	2.2
AGAVCC	.3640	.2075	1.8
AMMOT	.9894	.2067	4.8
EEMECH	.5970	.2107	2.8
IWPRPR	.8314	.3294	2.5
COSPEC	.4656	.1365	3.4
AGAZ	1.0887	.4731	2.3
AGMAROC	.2650	.1408	1.9
(W)SEADU	.4098	.1739	2.4
(E)SEADU	.6699	.2222	3.0
CBTENG	.5091	.2614	1.9
FARTYFC	.6064	.1802	3.4
AGAV	5.5397	2.3412	2.4
AGAVR	7.1142	2.6134	2.7
BECF03	2.9298	.5522	5.3
BECF10	1.7401	.4715	3.7
HQBAD	.5822	.0994	5.9
AGMARAK	.3993	.1332	3.0
AUTOMEC	.4408	.1153	3.8
AGADJ	.9310	.1624	5.7
AGBHEL	1.2029	.0826	14.6
PIBAD	.1727	.0768	2.2
COMMCTR	3.2016	.1876	17.1
FRADIO	1.0343	.2680	3.9
AGAFAM	.3292	.0472	7.0

TABLE 3

NAMES AND ABBREVIATIONS FOR
26 MARINE CORPS SCHOOLS

Abbreviation	Title
FOODSER	Basic Food Service Course
MP	Law Enforcement (MP) Course
AGAVCC	Aviation Crash Crewman Course
AMMOT	Ammunition Storage Course
EEMECH	Basic Engineers Equipment Mechanics Course
IWPRPR	Small Arms Repair Course
COSPEC	Law Enforcement (Corrections Specialist) Course
AGAZ	Aviation Maintenance Administration
AGMAROC	Marine Aviation Operations Clerical
(W)SEADU	Sea Duty Indoctrination Course (West Coast)
(E)SEADU	Sea Duty Indoctrination Course (East Coast)
CBTENG	Basic Combat Engineer Course
FARTYFC	Field Artillery Fire Control Course
AGAV	Avionics Technician Course
AGAVR	Avionics Repair Course
BECF03	Radio Fundamentals Course
BECF10	Ground Radio Repair Course
HQBAD	Basic Administration Course (Camp Pendleton)
AGMARAK	Marine Aviation Supply (Mechanical)
AUTOMEC	Basic Auto Mechanics Course
AGADJ	Aviation Machinist Mate (Jet Engine) Course
AGBHEL	Basic Helicopter Course
PIBAD	Basic Administration Course (Parris Island)
COMMCTR	Communications Center Man Course
FRADIO	Field Radio Operator Course
AGAFAM	Aviation Familiarization Course

TABLE 4

MATRIX OF CORRELATIONS FROM FY 75 60,000-MAN SAMPLE

	VE	AR	PA	CI	MA	ACS	ARC	GIT	SM	AI	ELI
VE	1.0000	0.6564	0.4913	0.5029	0.5677	0.4188	0.4119	0.6905	0.5859	0.4521	0.4965
AR	0.6564	1.0000	0.5683	0.4396	0.5266	0.5186	0.4408	0.5843	0.5339	0.4276	0.4450
PA	0.4913	0.5683	1.0000	0.3520	0.5082	0.4600	0.3942	0.4746	0.5048	0.4286	0.4776
CI	0.5029	0.4396	0.3520	1.0000	0.4555	0.4190	0.2927	0.5057	0.4507	0.3645	0.3318
MA	0.5677	0.5266	0.5082	0.4555	1.0000	0.4516	0.3845	0.6006	0.6643	0.5871	0.5115
ACS	-0.4188	0.5186	0.4600	0.4190	0.4516	1.0000	0.4591	0.4292	0.4359	0.3283	0.2927
ARC	0.4119	0.4408	0.3942	0.2927	0.3845	0.4591	1.0000	0.3722	0.3531	0.2474	0.2550
GIT	0.6905	0.5843	0.4746	0.5057	0.6006	0.4292	0.3722	1.0000	0.6330	0.5896	0.5270
SM	0.5859	0.5339	0.5048	0.4507	0.6643	0.4359	0.3531	0.6330	1.0000	0.6728	0.5576
AI	0.4521	0.4276	0.4286	0.3645	0.5871	0.3283	0.2474	0.5896	0.6728	1.0000	0.5723
ELI	0.4965	0.4450	0.4776	0.3318	0.5115	0.2927	0.2550	0.5270	0.5576	0.5723	1.0000

TABLE 5

MATRIX OF UNCORRECTED CORRELATION COEFFICIENTS

	VE	AR	PA	CI	MA	ACS	ARC	GIT	SM	AI	ELI
VE	1.0000	0.5022	0.3348	0.3231	0.4270	0.2487	0.2581	0.5580	0.4465	0.2905	0.4194
AP	0.5022	1.0000	0.4323	0.2633	0.4049	0.3832	0.2994	0.4187	0.3920	0.2670	0.3159
PA	0.3348	0.4323	1.0000	0.1727	0.4225	0.3233	0.2862	0.3241	0.4018	0.5135	0.3578
CI	0.3231	0.2633	0.1727	1.0000	0.3207	0.1924	0.1365	0.3510	0.2893	0.2341	0.2046
MA	0.4270	0.4049	0.4225	0.3207	1.0000	0.3184	0.2900	0.4746	0.5324	0.4556	0.4136
ACS	0.2487	0.3832	0.3233	0.1924	0.3184	1.0000	0.3454	0.2459	0.2843	0.1528	0.1690
ARC	0.2581	0.2994	0.2862	0.1365	0.2900	0.3454	1.0000	0.2250	0.2410	0.1163	0.1350
GIT	0.5580	0.4187	0.3241	0.3510	0.4746	0.2459	0.2250	1.0000	0.5464	0.4997	0.4468
SM	0.4465	0.3920	0.4018	0.2893	0.5324	0.2843	0.2410	0.5464	1.0000	0.5836	0.4545
AI	0.2905	0.2670	0.5135	0.2341	0.4556	0.1528	0.1163	0.4997	0.5836	1.0000	0.4804
ELI	0.4194	0.3169	0.3578	0.2046	0.4136	0.1690	0.1350	0.4468	0.4545	0.4804	1.0000

TABLE 6

MATRIX CORRECTED FOR MULTIPLE CURTAILMENT

	VE	AR	PA	CI	MA	ACS	ARC	GIT	SM	AI	ELI
VE	1.0000	0.6564	0.4913	0.5029	0.5677	0.4188	0.4119	0.6876	0.5886	0.4102	0.5450
AR	0.6564	1.0000	0.5683	0.4396	0.5266	0.5186	0.4408	0.5690	0.5352	0.3806	0.4546
PA	0.4913	0.5683	1.0000	0.3520	0.5082	0.4600	0.3942	0.4634	0.5199	0.4061	0.4692
CI	0.5029	0.4396	0.3520	1.0000	0.4555	0.4190	0.2927	0.4991	0.4409	0.3424	0.3445
MA	0.5677	0.5266	0.5082	0.4555	1.0000	0.4516	0.3845	0.5824	0.6186	0.5218	0.5083
ACS	0.4188	0.5186	0.4600	0.4190	0.4516	1.0000	0.4591	0.4068	0.4292	0.2735	0.3057
ARC	0.4119	0.4408	0.3942	0.2927	0.3845	0.4591	1.0000	0.3607	0.3618	0.2136	0.2553
GIT	0.6876	0.5690	0.4634	0.4991	0.5824	0.4068	0.3607	1.0000	0.6461	0.5672	0.5507
SM	0.5886	0.5352	0.5199	0.4409	0.6186	0.4292	0.3618	0.6461	1.0000	0.6365	0.5501
AI	0.4102	0.3806	0.4061	0.3424	0.5218	0.2735	0.2136	0.5672	0.6365	1.0000	0.5434
ELI	0.5450	0.4546	0.4692	0.3445	0.5083	0.3057	0.2553	0.5507	0.5501	0.5434	1.0000

TABLE 7

DIFFERENCE BETWEEN 60,000-MAN MATRIX AND
MATRIX CORRECTED FOR MULTIPLE CURTAILMENT

	VE	AR	PA	CI	MA	ACS	ARC	GIT	SM	AI	ELI
VE	-0.0000	0.0000	-0.0000	0.0000	0.0000	0.0000	0.0000	0.0028	-0.0026	0.0420	-0.0486
AR	0.0000	0.0000	0.0000	-0.0000	0.0000	0.0000	0.0000	0.0152	-0.0012	0.0471	-0.0095
PA	0.0000	0.0000	-0.0000	0.0000	-0.0000	0.0000	0.0000	0.0111	-0.0151	0.0224	0.0084
CI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0066	0.0098	0.0221	-0.0127
MA	0.0000	0.0000	0.0000	-0.0000	0.0000	-0.0000	0.0000	0.0182	0.0458	0.0653	0.0031
ACS	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0000	0.0000	0.0225	0.0068	0.0549	-0.0130
ARC	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0000	0.0000	0.0114	-0.0087	0.0339	-0.0003
GIT	0.0028	0.0152	0.0111	0.0066	0.0182	0.0225	0.0114	-0.0000	-0.0131	0.0223	-0.0237
SM	-0.0026	-0.0012	-0.0151	0.0098	0.0458	0.0068	-0.0087	-0.0131	-0.0000	0.0363	0.0075
AI	0.0420	0.0471	0.0224	0.0221	0.0653	0.0549	0.0339	0.0223	0.0363	-0.0000	0.0289
ELI	-0.0486	-0.0095	0.0084	-0.0127	0.0031	-0.0130	-0.0003	-0.0237	0.0075	0.0289	0.0000

TABLE 8

MATRIX OF CORRELATION COEFFICIENTS ASSUMING DIRECT CURTAILMENT ON VE

	VE	AR	PA	CI	MA	ACS	ARC	GIT	SM	AI	ELI
VE	1.0000	0.6072	0.4236	0.4098	0.5279	0.3201	0.3317	0.6628	0.5490	0.3711	0.5195
AR	0.6072	1.0000	0.4905	0.3383	0.4849	0.4265	0.3537	0.5172	0.4773	0.3333	0.4074
PA	0.4236	0.4905	1.0000	0.2334	0.4750	0.3613	0.3280	0.3998	0.4591	0.3589	0.4168
CI	0.4098	0.3383	0.2334	1.0000	0.3818	0.2368	0.1859	0.4201	0.3555	0.2833	0.2756
MA	0.5279	0.4849	0.4750	0.3818	1.0000	0.3639	0.3400	0.5501	0.5896	0.4981	0.4815
ACS	0.3201	0.4265	0.3613	0.2368	0.3639	1.0000	0.3748	0.3067	0.3341	0.1953	0.2259
ARC	0.3317	0.3537	0.3280	0.1859	0.3400	0.3748	1.0000	0.2912	0.2968	0.1623	0.1969
GIT	0.6628	0.5172	0.3998	0.4201	0.5501	0.3067	0.2912	1.0000	0.6143	0.5416	0.5250
SM	0.5490	0.4773	0.4591	0.3555	0.5896	0.3341	0.2968	0.6143	1.0000	0.6151	0.5202
AI	0.3711	0.3333	0.3589	0.2833	0.4981	0.1953	0.1623	0.5416	0.6151	1.0000	0.5203
ELI	0.5195	0.4074	0.4168	0.2756	0.4815	0.2259	0.1969	0.5250	0.5202	0.5203	1.0000

TABLE 9

DIFFERENCE BETWEEN 60,000-MAN MATRIX AND MATRIX CORRECTED FOR CURTAILMENT ON VE

	VE	AR	PA	CI	MA	ACS	ARC	GIT	SM	AI	ELI
VE	0.0000	0.0491	0.0676	0.0931	0.0398	0.0987	0.0802	0.0277	0.0370	0.0810	-0.0231
AR	0.0491	0.0000	0.0778	0.1013	0.0417	0.0922	0.0871	0.0670	0.0566	0.0944	0.0376
PA	0.0676	0.0778	-0.0000	0.1186	0.0322	0.0987	0.0662	0.0748	0.0457	0.0697	0.0608
CI	0.0931	0.1013	0.1186	0.0000	0.0738	0.1822	0.1068	0.0856	0.0952	0.0812	0.0562
MA	0.0398	0.0417	0.0322	0.0738	0.0000	0.0876	0.0445	0.0505	0.0747	0.0890	0.0299
ACS	0.0987	0.0922	0.0987	0.1822	0.0876	-0.0000	0.0844	0.1225	0.1019	0.1331	0.0669
ARC	0.0802	0.0871	0.0662	0.1068	0.0445	0.0844	0.0000	0.0810	0.0563	0.0852	0.0581
GIT	0.0277	0.0670	0.0748	0.0856	0.0505	0.1225	0.0810	0.0000	0.0186	0.0480	0.0020
SM	0.0370	0.0566	0.0457	0.0952	0.0747	0.1019	0.0563	0.0186	0.0000	0.0576	0.0374
AI	0.0810	0.0944	0.0697	0.0812	0.0890	0.1331	0.0852	0.0480	0.0576	0.0000	0.0520
ELI	-0.0231	0.0376	0.0608	0.0562	0.0299	0.0669	0.0581	0.0020	0.0374	0.0520	-0.0000

-22-

REFERENCES

1. Center for Naval Analyses, Study 1084, "An Analysis of Marine Corps School Assignment and Performance," by Steve Verna and Thomas L. Mifflin, forthcoming

2. Pearson, Karl, "On the Influence of Natural Selection on the Variability and Correlation of Organs," Phil. Trans. Roy. Soc. A, pp. 1-66

3. Burt, Cyril, "Validation Tests for Personnel Selection," British Journal of Psychology, 1943, 34, pp. 1-19

4. Lawly, D.N., "A Note on Karl Pearson's Selection Formulae," Proc. Royal Soc. Edinburgh, Sec. A, 1943, 62 Part I, pp. 28-30

5. Gulliksen, Harold, Theory of Mental Tests, John Wiley & Sons, New York, N.Y., pp. 145-171

6. Thorndike, Robert L., Personnel Selection, John Wiley & Sons, New York, N.Y., pp. 169-176

APPENDIX A

FORTRAN PROGRAM

APPENDIX A

FORTRAN PROGRAM

This appendix is the program used to correct the effects of multiple curtailment on the correlation matrixes of 84 Marine Corps schools.

The notation used in the program is slightly different from the notation used in this report. For example, $\hat{D}$ and $\hat{C}$ are denoted by DPRIME and CPRIME. Also, instead of using "t" as a subscript, "x" is used; for example, the matrix DXA in the program corresponds to D_{ta} in the report. Finally, since the C_{aa} matrix is inverted, we used the IBM routine IMINV.

```
      PROGRAM RANGE
      DIMENSION LM(11),MM(11)
      REAL OSIGMA(27),OPRIME(20,20),CSIGMA(27),CPRIME(27,27),CXX(15,16),
     + C(27,27),OAA(20,20),HOLD(11,11),CAA(11,11),CXA(15,11),CMEAN(27),
     +CAX(11,16),DXA(15,11),DAX(11,16),              S(11,11),
     +VXA(16,11),OMINUSC(11,16),OANSWER(27,27),OCORR(27,27),OXX(16,16)
      INTEGER AA,T
C
C
C
C     READ IN 'AA' THE NUMBER OF DIRECTLY CURTAILED VARIABLES
C
      READ 987,AA
C
C
C
C
C     READ IN 'OSIGMA' THE STANDARD DEVIATIONS OF THE DIRECTLY CURTAILED
C     READ IN 'OPRIME' THE MATRIX OF CORRELATIONS OF THE DIRECTLY
C          CURTAILED VARIABLES IN THE GENERAL POPULATION
C
      READ 11,(OSIGMA(I),I=1,20)
      DO 782 I=1,20
      READ 542,(OPRIME(I,J),J=1,20)
C
C
C
C     CONVERT OPRIME INTO 'OAA' THE VARIANCE-COVARIANCE MATRIX
C
      DO 783 J=1,20
  783 OAA(I,J)=OPRIME(I,J)*OSIGMA(I)*OSIGMA(J)
  782 CONTINUE
C
C
C
C     READ IN 'T' THE NUMBER OF INDIRECTLY CURTAILED VARIABLES, THE NAME
C          OF THIS SCHOOL, AND THE NUMBER OF MEN IN IT
C
    1 READ 1005, T,NAME1,NAME2
      IF (T.EQ.0) GO TO 9999
      READ 100, XNOMEN
      L=T+AA
      PRINT 1004, NAME1,NAME2,XNOMEN,L
C
C
C
C     READ IN THE MEANS, STANDARD DEVIATIONS, AND THE MATRIX OF CORRELATIONS
C          FOR ALL THE VARIABLES (BOTH CURTAILED AND UNCURTAILED) IN THIS SCHOOL
C
      READ 11, (CMEAN(I),I=1,L)
      READ 11,(CSIGMA(I),I=1,L)
      DO 924 I=1,L
  924 READ 11,(CPRIME(I,J),J=1,L)
C
```

A-2

```
C
C
C     CONVERT CPRIME INTO VARIANCE-COVARIANCE MATRIX,  'C'
C
      DO 81,J=1,L
      DO 81,I=1,L
  81  C(I,J)=CPRIME(I,J)*CSIGMA(I)*CSIGMA(J)
C
C
C
C     SPLIT C INTO ITS COMPONENT MATRICES,  'CAA', 'CXX', 'CXA', 'CAX'
C
      DO 113 J=1,AA
      DO 113 I=1,AA
      CAA(I,J)=C(I,J)
  113 HOLD(I,J)=CAA(I,J)
C
      DO 30 J=1,T
      L=AA+J
      DO 30 I=1,T
      K=AA+I
  30  CXX(I,J)=C(K,L)
C
      DO 40 I=1,T
      K=AA+I
      DO 40 J=1,AA
      CXA(I,J)=C(K,J)
  40  CAX(J,I)=CXA(I,J)
C
C
C
C   CALCULATE  'DXA'  AND ITS TRANSPOSE, 'DAX'
C
C   FIRST INVERT CAA.  PRINT OUT THE DETERMINANT OF CAA, AND THE PRODUCT
C       OF CAA AND ITS INVERSE TO BE SURE CAA IS NONSINGULAR
C
      CALL IMINV(CAA,AA,DET,LM,MM)
      PRINT 457,DET
      IF(DET.EQ.0) PRINT 129
      DO 1001 I=1,AA
      DO 1001, J=1,AA
      SUM=0
      DO 1002 K=1,AA
 1002 SUM=HOLD(I,K)*CAA(K,J)+SUM
 1001 S(I,J)=SUM
      DO 925 I=1,AA
 925  PRINT 926,(S(I,J),J=1,AA)
      PRINT 1007
C
C     NOW CALCULATE DXA
C
      DO 78 I=1,T
      DO 78 J=1,AA
      SUM=0
      DO 178 K=1,AA
 178  SUM=CXA(I,K)*CAA(K,J)+SUM
 78   VXA(I,J)=SUM
C
      DO 103 I=1,T
```

BEST AVAILABLE COPY

A-3

```
          DO 103 J=1,AA
          SUM=0
          CO 104 K=1,AA
 104      SUM=VXA(I,K)*OAA(K,J)+SUM
          OXA(I,J)=SUM
 103   OAX(J,I)=DXA(I,J)
C
C
C
C     CALCULATE OXX
C
C     FIRST COMPUTE  'DMINUSC' =OAX-CAX
C
       DO 79 J=1,T
       CO 79 I=1,AA
 79    CMINUSC(I,J)=OAX(I,J)-CAX(I,J)
C
       CO 83,J=1,T
       DO 83 I=1,T
       SUM=0
       CO 166 K=1,AA
 166   SUM=VXA(I,K)*DMINUSC(K,J)+SUM
 83    DXX(I,J)=CXX(I,J)+SUM
C
C
C
C     CCNSTRUCT  'OANSWER'  FROM THE FOUR SUBMATRICES OAA, OXA, OAX, OXX
C
       DO 92 I=1,T
       L=I+AA
       CO 92 J=1,AA
       DANSWER(L,J)=OXA(I,J)
 92    DANSWEP(J,L)=CANSWER(L,J)
C
       DO 93 J=1,T
       M=AA+J
       CO 93 I=1,T
       L=AA+I
 93    CANSWER(L,M)=DXX(I,J)
C
       DO 91 J=1,20
       DO 91 I=1,20
 91    CANSWER(I,J)=DAA(I,J)
C
C
C
C     COMPUTE THE STANDARD DEVIATIONS OF THE INDIRECTLY CURTAILED VARIABLES
C
       L=T+AA
       DO 95 I=21,L
 95    CSIGMA(I)=SQRT(DANSWER(I,I))
C
C
C
C     CONVERT DANSWER INTO THE CORRELATION MATRIX, 'OCORR'
C
       CO 94 J=1,L
       DO 94 I=1,L
 94    OCORR(I,J)=DANSWER(I,J)/(DSIGMA(I)*DSIGMA(J))
```

A-4

```
C
C
C
C      OUTPUT THE MEANS, STANDARD DEVIATIONS AND CORRECTED CORRELATION MATRIX
C
       PRINT 1008,(CMEAN(I),I=1,L)
       WRITE( 1,12),(CMEAN(I),I=1,L)
       PRINT 1008,(DSIGMA(I),I=1,L)
       WRITE( 1,12) ,(DSIGMA(I),I=1,L)
       DO 927 I=1,L
       WRITE( 1,11) ,(DCORR(I,J),J=1,L)
  927  PRINT 1100, (DCORR(I,J),J=1,L)
       ENDFILE 1
       GO TO 1
C
   11    FORMAT(8F10.7)
    12 FORMAT(8F10.6)
   100 FORMAT(F4.0)
   129   FORMAT(1H ,*SINGULARITY*)
   457   FORMAT(1H0,*THE DET=*,E14.7)
   542   FORMAT(8F10.7/8F10.7/4F10.7)
    926 FORMAT(1X,20F6.3)
   987   FORMAT(I2)
  1004 FORMAT(1H1,* THIS IS SCHOOL *,2A7,* WITH *,F4.0,* MEN*,//,
      1* L IS *,I5)
  1005 FORMAT(I2,2A7)
  1006 FORMAT(A7)
  1007 FORMAT(///)
  1008 FORMAT(1X,10F12.7)
  1100 FORMAT(1X,10F10.7)
  9999 STOP
       END
          SCOPE
W=  00  LOAD
```

A-5

APPENDIX B

APL PROGRAM

APPENDIX B

APL PROGRAM

This appendix contains an APL version of the computer program used to correct for multiple curtailment. The definitions of variables are given below; the names in parentheses are the names of the corresponding variables as given in the text.

ALLCOR — matrix of correlations of directly restricted variables in the general population ($\hat{D}_{aa}$). Must be supplied.

CLASSCOR — matrix of correlations of all variables in restricted population ($\hat{C}$). Must be supplied.

ASIZE — number of directly curtailed variables (a).

XSIZE — number of indirectly curtailed variables (t).

C — variance-covariance matrix computed from CLASSCOR (C).

CAA, CXA, CXX — variance-covariance submatrixes of C (C_{aa}, C_{ta}, C_{tt}).

DAA — variance-covariance matrix computed from ALLCOR (D_{aa}).

DXA, DXX — variance-covariance matrixes computed by the program (D_{ta}, D_{tt}).

ANSWXA, ANSWXX — correlation matrixes computed from DXA and DXX ($\hat{D}_{ta}$, $\hat{D}_{tt}$).

SIGC, SIGD — vectors of standard deviations in restricted and general populations, ($\{\sigma^r\}$, $\{\sigma^g\}$), respectively.

The actual correction equations are found in statements 28 through 31. The rest of the program initializes the arrays and converts to and from correlation and variance-covariance matrixes.

```
      ∇ CLASSCOR CORRECTION ALLCOR
[1]     ASIZE←1↑(ρALLCOR)
[2]     XSIZE←1↑((ρCLASSCOR)-ρALLCOR)
[3]     TOTAL←ASIZE+XSIZE
[4]     C←(TOTAL,TOTAL)ρ0
[5]     DAA←(ASIZE,ASIZE)ρ0
[6]     ANSWXX←(XSIZE,XSIZE)ρ0
[7]     ANSWXA←(XSIZE,ASIZE)ρ0
[8]     ⍝CONVERT TO VAR-COVAR MATRICES
[9]     I←1
[10] ROW:J←I
[11] COL:C[I;J]←C[J;I]←CLASSCOR[I;J]×SIGC[I]×SIGC[J]
[12]    J←J+1
[13]    →COL×ι(J≤TOTAL)
[14]    I←I+1
[15]    →ROW×ι(I≤TOTAL)
[16]    CAA←(ASIZE,ASIZE)↑C
[17]    CAX←(ASIZE,-XSIZE)↑C
[18]    CXA←⍉CAX
[19]    CXX←(-XSIZE,XSIZE)↑C
[20]    ⍝FORM DAA
[21]    I←1
[22] ROW1:J←I
[23] COL1:DAA[I;J]←DAA[J;I]←ALLCOR[I;J]×SIGD[I]×SIGD[J]
[24]    J←J+1
[25]    →COL1×ι(J≤ASIZE)
[26]    I←I+1
[27]    →ROW1×ι(I≤ASIZE)
[28]    VXA←CXA+.×(⍉CAA)
[29]    DXA←VXA+.×DAA
[30]    DMCAX←(⍉DXA)-CAX
[31]    DXX←CXX+(VXA+.×DMCAX)
[32]    ⍝CONVERT DXA AND DXX BACK TO CORRELATION MATRICES
[33]    SIGD[TOTAL]←DXX[XSIZE;XSIZE]*0.5
[34]    I←1
[35] ROW2:J←I
[36] COL2:ANSWXX[I;J]←ANSWXX[J;I]←DXX[I;J]÷(SIGD[ASIZE+I]×SIGD[ASIZE+J])
[37]    J←J+1
[38]    →COL2×ι(J≤XSIZE)
[39]    I←I+1
[40]    →ROW2×ι(I≤XSIZE)
[41]    ANSWXX[XSIZE;]←ANSWXX[;XSIZE]
[42]    ⍝NOW DO DXA
[43]    I←1
[44] ROW3:J←1
[45] COL3:ANSWXA[I;J]←DXA[I;J]÷(SIGD[ASIZE+I]×SIGD[J])
[46]    J←J+1
[47]    →COL3×ι(J≤ASIZE)
[48]    I←I+1
[49]    →ROW3×ι(I≤XSIZE)
[50]    '          THE RESULTING CORRECTED CORRELATION MATRIX IS'
[51]    ⎕ ←(ALLCOR,[1]ANSWXA),((⍉ANSWXA),[1]ANSWXX)
      ∇
```

APPENDIX C

EQUALITY OF PARTIAL CORRELATIONS

APPENDIX C

EQUALITY OF PARTIAL CORRELATIONS

One of the key assumptions in deriving the range correction equations is that after the effects of the explicitly selected variables are partialled out, the correlations between the variables subject to incidental selection in the restricted population are the same as the analogous correlations in the uncurtailed population. This appendix indicates how this assumption is valid in the example of the Marine training schools.

Just as in the COMPARISON WITH ANOTHER TECHNIQUE section of this report, the first seven ACB-61 subtests were arbitrarily designated as the directly restricted variables for four Marine training schools. Table C-1 shows the partial correlations between the other four variables (i.e., the incidentally selected variables) controlling for the effects of the directly restricted variables in both the school population and that of all FY-1975 Marine recruits.

The similarity of the first four columns of table C-1 and the last column indicates that this key assumption is justified when several variables are the basis of curtailment. Similarly, table C-2 shows the accuracy of the assumption that is used when only one variable is the basis for curtailment.

TABLE C-1

PARTIAL CORRELATIONS BETWEEN THE LAST FOUR ACB-61 SUBTESTS CONTROLLING FOR THE FIRST SEVEN SUBTESTS

Variable pair	School				All Marine recruits in FY 1975
	AGAFAM	FRADIO	COMMCTR	PIBAD	
ELI with AI	.32	.33	.31	.37	.34
ELI with SM	.22	.24	.20	.27	.24
ELI with GIT	.20	.08	.13	.20	.18
AI with SM	.43	.44	.45	.44	.42
AI with GIT	.34	.32	.39	.29	.31
SM with GIT	.29	.24	.28	.26	.23

TABLE C-2

PARTIAL CORRELATIONS BETWEEN THE LAST FOUR ACB-61
SUBTESTS CONTROLLING FOR THE FIRST SUBTEST

Variable pair	AGAFAM	School FRADIO	COMMCTR	PIBAD	All Marine recruits in FY 1975
ELI with AI	.41	.37	.42	.55	.45
ELI with SM	.33	.29	.35	.53	.38
ELI with GIT	.28	.13	.24	.47	.29
AI with SM	.53	.53	.55	.64	.56
AI with GIT	.43	.39	.47	.52	.43
SM with GIT	.40	.35	.39	.61	.39